Can We Test Our Way Out of the COVID-19 Pandemic?

Matthew A. Pettengill\textsuperscript{a} and Alexander J. McAdam\textsuperscript{b}

\textsuperscript{a}. Department of Pathology, Anatomy, and Cell Biology, Thomas Jefferson University, Philadelphia, Pennsylvania, USA.

\textsuperscript{b}. Department of Laboratory Medicine, Boston Children’s Hospital, Boston Massachusetts, USA.

Alexander.mcadam@childrens.harvard.edu

Running Head: Testing our way out of a pandemic?

#Address Correspondence to Matthew A. Pettengill, Matthew.Pettengill@jefferson.edu
Abstract: Frequent, low cost, universal testing for SARS-CoV-2 with quarantine of those with a positive result has been suggested as a strategy to address the COVID-19 pandemic in the United States. Specifically, home or community use of test that use paper-strip detection devices, which may have reduced sensitivity for SARS-CoV-2, has been advocated. There are several potential challenges or problems with this strategy, including the availability of such tests, consequences of incorrect test results, difficulties with adherence to testing, and the accuracy of such tests for detection of infectious people. Because of these, we think it is premature to strongly advocate for such a testing strategy, as the adverse consequences may outweigh any benefits. High-quality outcomes data demonstrating the efficacy of this testing strategy are needed before widespread implementation.

Recently, the use of universal, frequent testing for COVID-19 to guide quarantine has been proposed and widely discussed as a method to dramatically reduce or eliminate community spread of SARS-CoV-2 (1). The core of the proposal is the mass production and utilization of inexpensive, rapid paper strip-based tests that can be used frequently (e.g. daily) in the community, whether in the home, workplace or school. The proposal is that if such results were available daily to everyone (or most people) in the country, transmission could be reduced enough to suppress the pandemic. We have several concerns about this approach, both technical and practical.

The current tests for acute SARS-CoV-2 infection in the US include mainly commercial or laboratory developed nucleic acid amplification tests (NAATs, such as PCR) with excellent technical sensitivity but somewhat lower clinical sensitivity (2), and a smaller number of antigen based tests with lower technical and clinical sensitivity. Since early July 2020, more than 600,000 SARS-CoV-2 tests have been performed per day in the US, and currently the US has the 2nd highest per capita rate of testing in the world among countries for which data is available.
(https://coronavirus.jhu.edu/testing/international-comparison, accessed 20 Aug 2020) – and the highest number of positive results and deaths. Many countries have dramatically reduced community spread of COVID-19 without the scale of per capita testing currently available in the US. For example, Canada has less than half of the U.S. per capita testing capacity and dramatically lower per capita rates of transmission (see Johns Hopkins database cited above). Perhaps enhanced SARS-CoV-2 testing capacity in the absence of robust well-funded public health infrastructure and contact tracing capacity is similar to rapid molecular diagnostics for patients with sepsis that are not paired with antimicrobial stewardship intervention – compelling on paper but clinically ineffective (3). Unfortunately, results for NAATs for SARS-CoV-2 in the US currently have a long turn-around-time (TAT) in many laboratories, with some taking several days. Such delayed results are not useful for guiding decisions on quarantine to reduce transmission, and thus decrease the public health impact of the testing capacity in the US. Testing has been faster at hospital and public health labs, whereas many commercial reference laboratories have been slower as they are overwhelmed by testing volume from clinical facilities or testing sites without internal testing capacity. Large scale testing *alone* does not appear to have been the key to transmission control in some other countries.

**Will rapid, inexpensive strip-based tests be available in the United States?**

Paper strip-based tests which are simple, inexpensive, and appropriate for community use to detect SARS-CoV-2 are not yet available. There are currently three antigen tests with Emergency Use Authorization in the U.S (https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/vitro-diagnostics-euas#individual-antigen). None is approved for home use, although all can be used in patient care settings with a CLIA Certificate of Waiver. What about the tests that are in development? Tests in development at Sherlock Biosciences and Mammoth Biosciences that have been promoted as “paper-strip and other simple, daily COVID-19 tests” (1) use pre-amplification by reverse-transcriptase and loop mediated isothermal amplification
(LAMP, a NAAT) before detection by a lateral flow assay (4, 5). These two tests each take about an hour to run and require specific temperatures for reverse transcription and amplification, and so would require a testing device or platform. If these tests could be modified for use in the community, we cannot see how they would be available for $1 to $5, less than the cost of most home pregnancy or ovulation tests. Low (or no) cost testing would be particularly important, as access to testing has been unequal, with less testing availability in low-income and minority neighborhoods (https://www.sciencemag.org/news/2020/07/huge-hole-covid-19-testing-data-makes-it-harder-study-racial-disparities, accessed 8/24/2020). Such low cost for end-users could only be realized with tremendous financial subsidies, for example from the federal or state governments. Paper strip based tests targeting viral antigens have been developed as well, but details on their performance, cost and availability are lacking. It is critically important to understand real-world performance parameters of such tests before being able to evaluate any potential role in diagnostics or screening to reduce transmission. It is also an open question whether such tests can be developed and produced at the massive needed scale any more quickly than vaccines, which can be expected to dramatically reduce the spread of SARS-CoV-2.

What would be the impact of false positive results?

Studies in which models of test parameters have been used to evaluate effects on transmission of SARS-CoV-2 have paid little or no attention to the specificity of tests or the practical impacts of imperfect specificity. According to these models, both turn-around-time and test frequency are more important than test sensitivity for preventing transmission (6, 7). These models rely on theoretical test performance parameters and assume ideal test utilization and human behavior. In one study the impact of limit of detection (LOD, directly related to clinical sensitivity), test frequency and turn-around-time were compared for their relative impact on predicted transmission, with a viral load of $10^3$ copies of RNA/mL used to approximate molecular test LOD and $10^5$ copies of RNA/mL used for low-sensitivity
daily tests (LSDT) LOD (7). LOD was the least impactful of these parameters, and turn-around time and testing frequency were suggested to be the keys to surveillance leading to reduced transmission for COVID-19. Test specificity was explicitly excluded from consideration in this particular study, as it could not have had an impact given the study design; the automatons in the model accepted their fate as potential transmitters of disease and altruistically isolated themselves for the good of society. We agree that specificity is of reduced importance when the result for false positives is that people comply with quarantine.

In reality, if specificity is even modestly compromised, it will strike at the core of an important parameter for real world impact of testing: the reliability of the results. Some of the LSDTs under development are antigen-based tests. The few available studies have found SARS-CoV-2 antigen tests to have high specificities, but have generally included relatively small numbers of samples, making it difficult to detect small flaws in the specificity of these tests (8). The real-world specificity of antigen tests may be significantly lower than that suggested by studies performed by trained laboratory staff, as demonstrated by a large cluster of false positive antigen tests in Vermont (https://www.burlingtonfreepress.com/story/news/2020/07/23/covid-19-testing-how-versions-differ-speed-use-and-accuracy-coronavirus/5446176002/). We are unable to find specificity data for the assays that use LAMP and lateral-flow detection.

Antigen test performance has been evaluated in detail for another respiratory virus – influenza virus. A meta-analysis of 159 studies including 26 different antigen based influenza tests found pooled sensitivity of 62.3% (which is why many labs have abandoned this type of test for influenza) and pooled specificity of 98.2% (9). A recent publication on a newer influenza antigen test with improved sensitivity found specificity in a similar range with that found in the meta-analysis cited above, 98.4% (10). Specificity in the range of 98% is typically acceptable (even very good) for use in patients with high pre-test probability of having the infection, but when used on a massive scale among asymptomatic subjects
there are consequences for even modest reductions in specificity. For example, if LSDT specificity is 98% and tests are used by each of the ~325 million people in the U.S. every day, there would be a staggering 6.5 million false positive results each day. In low prevalence areas of the country, a very large proportion of positive tests would be false positive. If we consider following up all positives with a NAAT, in the example above there would be an order of magnitude more confirmatory tests per day required than current national capacity (approximately 650K tests per day, https://coronavirus.jhu.edu/testing/international-comparison, accessed 20Aug2020). How could we expect the general public to take the test results seriously and respond to them appropriately if the test is insensitive and many positive results are wrong?

Will people agree to frequent testing?

Another important consideration is whether a significant proportion of the population would be willing to use LSDTs, or to require their employees, clients, or students to use them. A meta-analysis of more than 100 studies found that physical distancing and wearing masks were associated with lower viral transmission (11). Both interventions are endorsed by the CDC and WHO, and yet their utilization has been politicized, and there is marked resistance to government mandates to wear masks in public spaces in many parts of the country. In locations where mask requirements exist, enforcement and compliance are variable. Similarly, regulations on large gatherings are inconsistently applied or non-existent. It is difficult to envision LSDTs gaining significant penetration into areas where transmission interventions are needed most, even if the results were reliable. If the sensitivity were low, as it is expected for antigen tests, and it is known that many positives are false positive, the results will not be generally believable and LSDTs would be unlikely to gain widespread utility. For example, if a test is 80% sensitive and 98% specific, the positive predictive value would be 67.7% and negative predictive value would be 98.9% at prevalence of 5%. If, however, the test described above were used at a massive scale in areas where prevalence is lower than 5%, which has likely been the case in most of the US during the
majority of time so far during the pandemic, PPV would suffer considerably: at 1% prevalence PPV would be 28.8%, and at 0.1% prevalence PPV would be 3.8%. It is also possible, perhaps likely, that people would use negative LSDT results (possibly falsely negative) to justify abandoning proven interventions, such as wearing a mask and social distancing.

We are additionally concerned that mass utilization of LSDT has the potential to negatively impact the public perception and trust of medical tests in general. There are already many in the US population that distrust some medical interventions (vaccines, for example) or existing tests (Lyme disease serology, for example). It is easy to envision a scenario with widespread daily use by untrained users that a substantial proportion of LSDT results for subjects who actually have the virus may be wrong, and that the majority of positive results will be wrong. If such a test is widely endorsed by the medical and laboratory communities but is ultimately considered by the public to be unreliable, it could lead to a dangerous erosion of the public trust in diagnostic tests in general. The ongoing saga of hydroxychloroquine as a potential therapy for COVID-19 may serve as a cautionary tale. There were promising in vitro data suggesting an anti-viral effect of the drug (12), and small studies in China and an uncontrolled, non-randomized trial in France were used by prominent physicians and scientists (and politicians) to suggest that hydroxychloroquine could help make the virus “disappear” (13). It didn’t (14).

Finally, if people do perform self-testing, some might decide not to report the results to their physician or public health authorities. This would compromise the quality of the data needed to track the pandemic. Furthermore, if asymptomatic people who have a positive result might not self-quarantine, perhaps because they do not believe the results or perhaps because they need to go to work, attend in-person classes or provide care to others.

Can insensitive tests accurately detect infectious individuals?
One argument in favor of frequent use of rapid tests is that the insensitivity of some of these tests, primarily antigen tests, is acceptable because the tests will still detect people who have high levels of virus and are therefore infectious (7). This raises two issues. First, do tests that could be used as LSDT specifically detect specimens with high viral loads? The data to address this are limited, but one antigen test was recently found to detect only 82% of nasopharyngeal samples with threshold cycle (Ct) values below 25 in a group of well-validated NAATs for SARS-CoV-2 (8). These low Ct values indicate that high levels of viral RNA were present, and so this antigen test does not reliably detect samples that are likely to contain high levels of SARS-CoV-2. Second, suppose that we hypothesize that less sensitive tests will reliably detect high levels of virus: does that mean that they will detect infectious individuals? Perhaps, but this is speculation. It is clear that samples with lower Ct values are more likely to contain SARS-CoV-2 that can be detected in viral culture (15, 16). But there are no data linking the Ct value or viral quantity to transmissibility, and it is possible that viral culture may be an insensitive indicator of the potential of an infected person to transmit SARS-CoV-2. Furthermore, there is no clear separation between samples that are infectious or not for viral culture using the Ct value, as there is significant overlap in the Ct value of samples that are positive and negative in viral culture (16). We do not dispute that people who have high levels of SARS-CoV-2 in their respiratory secretions are more likely to be infectious than those with low levels of virus, but we do question whether insensitive LSDTs can reliably detect whether someone is likely to be infectious.

Can we crush the curve without widespread testing?

Inexpensive, reasonably achievable measures such as use of masks, hand hygiene, staying home when ill and avoiding close contact are important in preventing transmission of SARS-CoV-2 (17-19) (https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/prevention.html). In this context, the primary role of testing should be to monitor the effects of these well-established practices, in addition to its obvious role in patient care of symptomatic individuals and infection control prior to...
medical procedures or hospital admission. But using testing to prevent transmission of SARS-CoV-2 on a large scale is like using the weather report to prevent global warming.

While the widespread, frequent use of LSDTs is appealing, and they may yet prove effective in select settings at reducing transmission, there is considerable cause for caution in extrapolating predicted performance and utility from models or results of antigen based testing for other respiratory viruses in the hands of testing professionals. We urge those promoting or considering LSDT strategies to demonstrate utility in a real world trial, and at a minimum to make lab-based performance characteristics publicly available, before prompting the general public to petition the US or state governments to approve and deploy LSDT.
References


